GWAS: REPORTS OF ITS DEATH HAVE BEEN GREATLY EXAGGERATED
Statistical Genetic Analysis in the Genomic Era

BY ALEXANDER WILSON, PH.D., NHGRI, AND JOAN BAILEY-WILSON, PH.D., NHGRI

The reports of the death of the genome-wide association study (GWAS) and the failure of the GWAS to find cures for common disease are eerily reminiscent of the reports of the death of linkage analysis and family studies a decade ago. Linkage studies identified the breast-cancer-associated genes *BRCA1* and *BRCA2* and other clinically relevant variants. For GWAS, time will tell. GWAS results are simply candidate regions, regions considerably narrower than those identified with linkage analysis, but candidate regions nevertheless. We need to do more work to identify the true causal variants underlying these associations and to understand their biology.

We believe there are several reasons why GWAS remains an important tool. First, it has already identified thousands of candidate variants across a variety of diseases and traits. With some exceptions (for example, macular degeneration), these variants are not yet clinically relevant but may lead to important discoveries in medical science by identifying new biological pathways and interactions among genes.

Second, more sophisticated analyses and meta-analyses that allow for dominance effects and epistatic and gene-environment interactions may yield a deeper understanding of the effects of variants on disease risk. Third, the GWAS has only recently been applied to traits other than disease risk, such as drug response. And fourth, genes identified with common variants may be useful in finding new drug targets and understanding adverse drug reactions, because the loci that harbor common variants of small effect may be important regulatory loci and/or may contain many rare variants with larger effects.

Rather than slavishly follow the method or tool du jour, we need to understand that different methods and study designs have the ability to detect different things, and that multiple designs and multiple tools must be used to understand complex traits. A GWAS is, in the broadest sense, simply a test of association between a trait and a set of genotypic markers that span the genome. A GWAS has several elements: the density of the markers, ranging from single-nucleotide polymorphism (SNP) panels to whole-genome sequence variants; the type of ascertainment (population- or family-based); the study design (case-control or family studies); and the type of trait.

The term GWAS, and the recent focus that has excluded most other methods and study designs, now refers to population-based studies of unrelated individuals, with high-density SNP panels in case-control study designs. This change in the meaning has been driven by the genotyping technology available, the ease of obtaining population-based samples of unrelated individuals, the focus on categorical disease, and the computational speed and simplicity of the analysis. This study design has good power to detect association with common variants, but then only common variants are being considered. However, other GWAS study designs are possible and have good power in other situations.

SNPs, copy-number variants, and rare sequence variants, collectively referred to here as sequence variants, have two attributes: the frequency of the alleles of the variant and the size of the effect of the variant alleles. Although much debate has focused on common versus rare alleles, what is important is the size of the effect of the variant alleles and whether the effect is at the individual, family, or population level.

One variant for familial hypercholesterolemia, for example, has a large effect in the individual with two copies of the variant, less of an effect in relatives who carry only one copy of the variant, and almost no effect in the population because the variant is so rare. Rare variants with very large effects can be difficult to detect in population studies, particularly if there is genetic heterogeneity (different variants that cause phenotypically indistinguishable disease, such as *BRCA1*). Population-based GWAS studies are not well-powered to detect such effects but a family-based or familial-case versus control GWAS can be more powerful in this situation.

When the cost of whole-genome sequencing drops to that of today's high-density SNP panels, we will have the ability to identify all variants. Family studies will be useful to enrich samples for specific rare variants and identify sporadic variants. Sampling designs similar to that used by the ClinSeq study may become the norm. In ClinSeq, a population-based sample is used to identify sequence variants, but consent includes contacting relatives of the individual in the population study to further study the effect of the variants. As the field of genetics advances, we should design our studies with the most appropriate tools for the job, not just the tool du jour. ●

The authors (who are married to each other) have both published in the areas of segregation analysis, linkage analysis, genome-wide association studies, and most recently analysis of sequence data. For more about the GWAS debate, see the article "Revolution Postponed" by Stephen S. Hall in the October 2010 issue of Scientific American. *(*Sci Amer ***303****:60–67, 2010; doi:10.1038/scientificamerican1010-60)*